Gabarito - Lista de exercícios 8

Cristiano de Carvalho Santos

Questão 1)

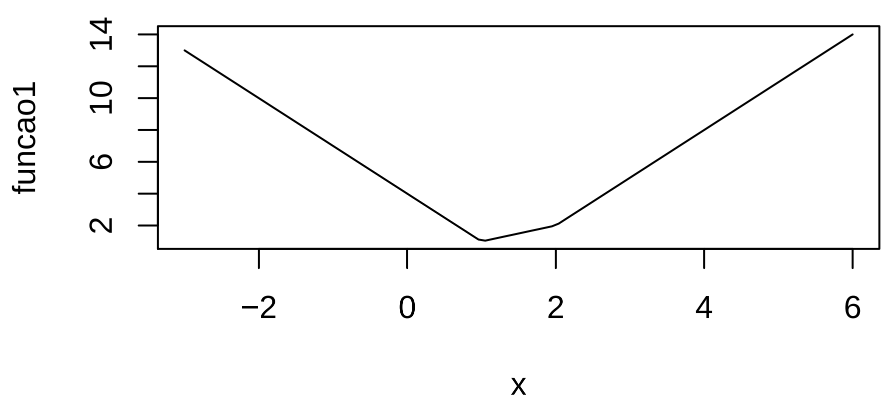
Minimize a função $f(x) = |x-2| + 2|x-1|$ e faça o gráfico da solução obtida.

```
xx <- c()
funcao1 <- function(x) {
  out <- abs(x-2) + 2*abs(x-1)
  xx <<- c(xx, x)
return(out)
}

out <- optimize(f = funcao1, interval = c(-3,3))
out
```

```
## $minimum
## [1] 1.000021
##
## $objective
## [1] 1.000021
```

```
#Grafico:
plot(funcao1, -3,6)
```



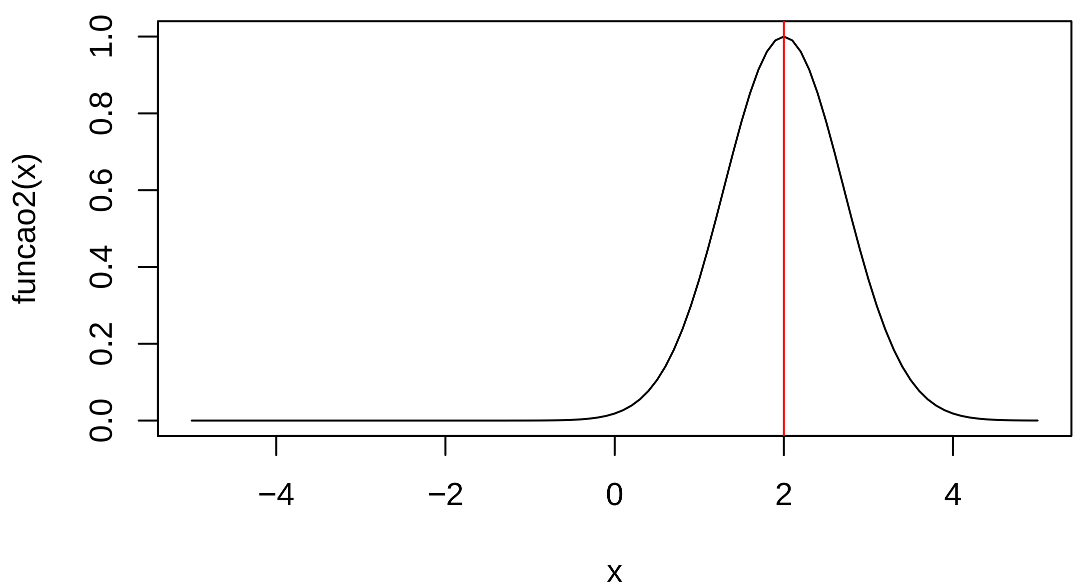
Questão 2)

Maximize a função $f(x) = e^{-(x-2)^2}$ e faça o gráfico da função com a solução obtida.

```
funcao2 <- function(x){
  exp(-( (x-2)^2 ))
}
out2<-optim(2, funcao2, method="CG")$par
out2
```

```
## [1] 2
```

```
#Grafico:
curve( funcao2(x), -5,5)
abline(v=out2, col=2)
```



Questão 3)

Encontre as raízes da função $sen(xcos(x))$ no intervalo de preferência e faça o gráfico da função com a solução obtida.

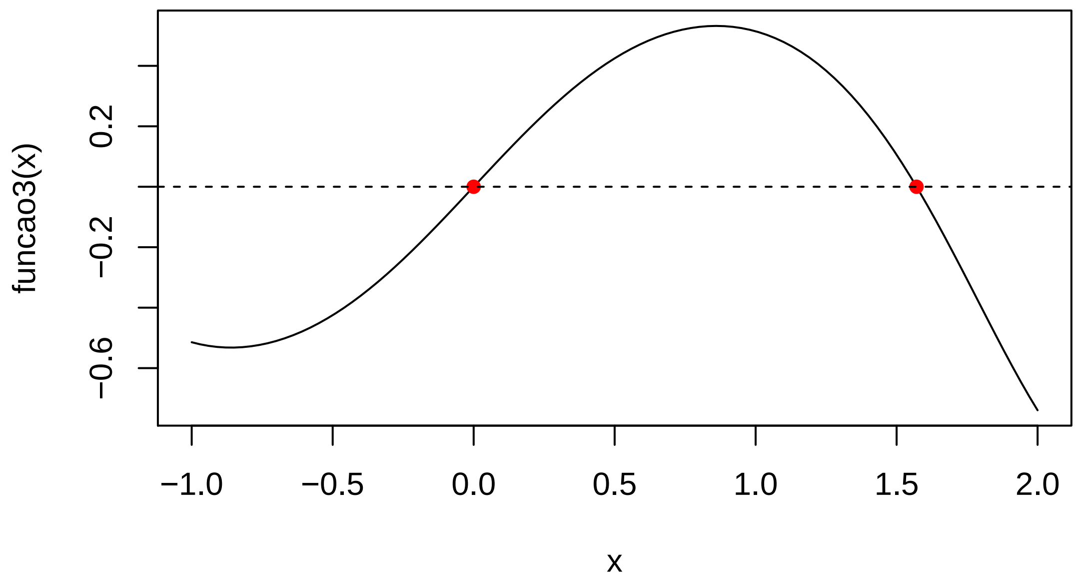
```
require(rootSolve)
```

```
## Loading required package: rootSolve
```

```
funcao3 <- function(x){
  sin(x*cos(x))
}

#Encontrando as raizes (por exemplo, no intervalo -1,2)
All <- uniroot.all(funcao3, c(-1, 2))
```

```r
#Grafico com os pontos:
curve(funcao3(x), -1,2)
points(All, y = rep(0, length(All)), pch = 16, cex = 1, col=2)
abline(h =0, lty= 2)
```



Questão 4)

A função $f(x) = (x^2 + y - 11)^2 + (x + y^2 - 7)^2$ é chamada de funcao Himmelblau, sendo esta muito utilizada para testar o desempenho de algoritmos de otimização. Para esta função, encontre os pontos de mínimo e máximo dentro do intervalo [-4,4], isto é, utilize esta informação para escolher os valores iniciais do algoritmo.

```r
funcao3_max <- function(x){
  -(x[1]^2 + x[2] - 11)^2 + (x[1] + x[2]^2 - 7)^2
}

funcao3_min <- function(x){
  (x[1]^2 + x[2] - 11)^2 + (x[1] + x[2]^2 - 7)^2
}
out3_max <- optim(par = c(0,0), funcao3_max)
out3_max$par
```

```
## [1] -0.08174154 -3.06213578
```

```r
out3_min <- optim(par = c(0,0), funcao3_min)
out3_min$par
```

```
## [1] 3.000098 1.999955
```

Questão 5)

O banco de dados "bad-drivers.txt" contém dados dos estados com os piores motoristas dos EUA. A reportagem por trás desses dados esta no seguinte link: https://fivethirtyeight.com/features/which-state-has-the-worst-drivers/. Utilize a variável "Number of drivers involved in fatal collisions per billion miles".

Dessa forma, encontre o intervalo de confiança de 95% para a média populacional e teste a hipótese de que a média populacional e igual 9,5 com 10% de significância.

```
#Carregando o banco de dados:
bad_drivers <- read.table("bad-drivers.txt", header=TRUE, sep=",")

### Utilizando a funcao t.test() para calculo do intervalo de confianca:
teste1 <- t.test(bad_drivers$Number.of.drivers.involved.in.fatal.collisions.per.billion.miles,
                 mu=15, conf.level = 0.95, alternative="two.sided")
teste1$conf.int
```

```
## [1] 14.63086 16.94953
## attr(,"conf.level")
## [1] 0.95
```

```
### Utilizando a funcao t.test() para calculo do teste de hipoteses:
teste2 <- t.test(bad_drivers$Number.of.drivers.involved.in.fatal.collisions.per.billion.miles,
                 mu=15, conf.level = 0.90, alternative="two.sided")
teste2 ## Note qeu nao eh preciso usar conf.level, visto que o pvalor eh fornecido.
```

```
##
##  One Sample t-test
##
## data:  bad_drivers$Number.of.drivers.involved.in.fatal.collisions.per.billion.miles
## t = 1.369, df = 50, p-value = 0.1771
## alternative hypothesis: true mean is not equal to 15
## 90 percent confidence interval:
##  14.82287 16.75752
## sample estimates:
## mean of x
##   15.7902
```

```
# Temos evidencias amostrais, com 10% de significancia, nao rejeitamos a hipotese
# que a media da amostra e igual a 15. Ou seja, a media de motoristas envolvidos
# em acidentes com colisoes que foram fatais e igual a 9,5.
```

Questão 6)

Um artigo da Nature(2003, Vol. 48, p.1013) descreveu um experimento para determinar o efeito de comer chocolate sobre uma medida de saúde cardiovascular para indivíduos que consumiram diferentes tipos de chocolate. Consideremos os resultados para somente para o tipo chocolate amargo e chocolate ao leite. No experimento, 12 indivíduos comeram 100 gramas de chocolate amargo e 200 gramas de chocolate ao leite e depois de uma hora a capacidade antioxidante total de seus plasmas sanguíneos foi medida em um ensaio. Os dados seguem abaixo:

chocolate amargo: 118.8, 122.6, 115.6, 113.6, 119.5, 115.9, 115.8, 115.1, 116.9, 115.4, 115.6, 107.9

chocolate ao leite: 102.1, 105.8, 99.6, 102.7, 98.8, 100.9, 102.8, 98.7, 94.7, 97.8, 99.7, 98.6

Inicialmente, teste se as variâncias populacionais são iguais. Utilizando a conclusão tirada no teste anterior, a um nível de 5%, com um teste para médias, verique se há evidências para sustentar a hipótese que consumir chocolate amargo produz um nível médio maior de capacidade antioxidante total do plasma sanguíneo quando comparado ao leite?

Para resolver esta questão, NÃO use a função t.test(), utilize o R como uma calculadora e faça os cálculos passo a passo.

```r
#Hipoteses:
#H0: mu1-m2 = 0
#H1: mu1-m2 > 0

# onde mu1 e a capacidade antioxidante media do plasma sanguineo resultante
# do consumo do chocolate amargo e mu2 e a capacidade antioxidante media do
# plasma sanguineo resultante do consumo do chocolate ao leite.

#Caso em que o sigma e desconhecido!
amargo <- c(118.8, 122.6, 115.6, 113.6, 119.5, 115.9, 115.8, 115.1, 116.9, 115.4, 115.6, 107.9)

leite <- c(102.1, 105.8, 99.6, 102.7, 98.8, 100.9, 102.8, 98.7, 94.7, 97.8, 99.7, 98.6)

#1) Testar se as variancias das duas amostas sao iguais ou nao.
n_amargo <- length(amargo)
n_leite <- length(leite)

var_amargo <- var(amargo)
var_leite <- var(leite)
estat_f <- var_amargo/var_leite

p_val <- 2 * pf(estat_f, n_amargo - 1, n_leite - 1, lower = F)
p_val
```

```
## [1] 0.5158203
```

```r
## ou com a função var.test
var.test(amargo, leite)
```

```
##
##  F test to compare two variances
##
## data:  amargo and leite
## F = 1.495, num df = 11, denom df = 11, p-value = 0.5158
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.4303878 5.1933105
## sample estimates:
## ratio of variances
##           1.495038
```

```r
# temos evidencias amostrais que as variancias sao iguais.
#===---===---===============---=====---===---===========



#2) Testar se a diferenca entre as medias é maior que zero.
diferenca_media <- mean(amargo) - mean(leite)

amostra_choc<- c(amargo, leite)

n_choc <- length(amargo) + length(leite)
```

```r
#calculo da variacia combinada para o caso das variancias iguais:
v_choc <- (((n_amargo - 1)*var(amargo)) + ((n_leite - 1)*var(leite))) / (n_amargo + n_leite - 2)


#Caso queira, pode-se calcular a estatistica T:
est_T <- (diferenca_media - 0) / sqrt(v_choc * ((1/n_amargo) + (1/n_leite)))
est_T
```

```
## [1] 12.04777
```

```r
#p-valor:
pt(est_T, n_amargo + n_leite - 2, lower = F)
```

```
## [1] 1.841695e-11
```

```r
#===---===---================---===---===---==========
## Apenas para comparacao
t.test(amargo, leite,alternative = "greater", var.equal = T  )
```

```
##
##  Two Sample t-test
##
## data:  amargo and leite
## t = 12.048, df = 22, p-value = 1.842e-11
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  13.61237      Inf
## sample estimates:
## mean of x mean of y
##  116.0583  100.1833
```

Questão 7)

Pensa-se que a concentração de um ingrediente ativo de um detergente líquido para lavagem de roupas seja afetada pelo tipo de catalizador empregado no processo. As concentrações estão descritas abaixo

Catalisador 1: 57.9, 66.2, 65.4, 65.4, 65.2, 62.6, 67.6, 63.7, 67.2, 71.0

Catalisador 2: 66.4, 71.7, 70.3, 69.3, 64.8, 69.6, 68.6, 69.4, 65.3, 68.6

Inicialmente, teste se as variâncias populacionais das concentrações dos catalisadores são iguais e, utilizando a conclusão tirada com este teste, obtenha o intervalo de confiança (5% de significância) para a diferenca das médias dos dois grupos.Para esta questão, você pode fazê-la utilizando o R como calculadora (como na questão 6) ou utilizando as funções var.test() e t.test().

```r
#Entrando com os dados:
Cat1 <- c(57.9, 66.2, 65.4, 65.4, 65.2, 62.6, 67.6, 63.7, 67.2, 71.0)
Cat2 <- c(66.4, 71.7, 70.3, 69.3, 64.8, 69.6, 68.6, 69.4, 65.3, 68.6)

### Primeiramente, efetuando o teste para verificar a igualdade das variancias:
var.test(Cat1,Cat2)
```

```
##
##  F test to compare two variances
##
## data:  Cat1 and Cat2
## F = 2.4049, num df = 9, denom df = 9, p-value = 0.2073
```

```
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.5973344 9.6819719
## sample estimates:
## ratio of variances
##           2.404865
```

```r
# temos evidencias amostrais que, com 5% de sigificancia, as variancias sao iguais
# Como as variancias sao iguais, mas desconhecidas, iremos efetuar o teste t
# considerando a igualdade entre as variancias.

### Agora, calculando o intervalo de confiança:
result <-t.test(Cat1, Cat2, var.equal = TRUE, conf.level = 0.95)
result$conf.int
```

```
## [1] -5.9028901 -0.4571099
## attr(,"conf.level")
## [1] 0.95
```

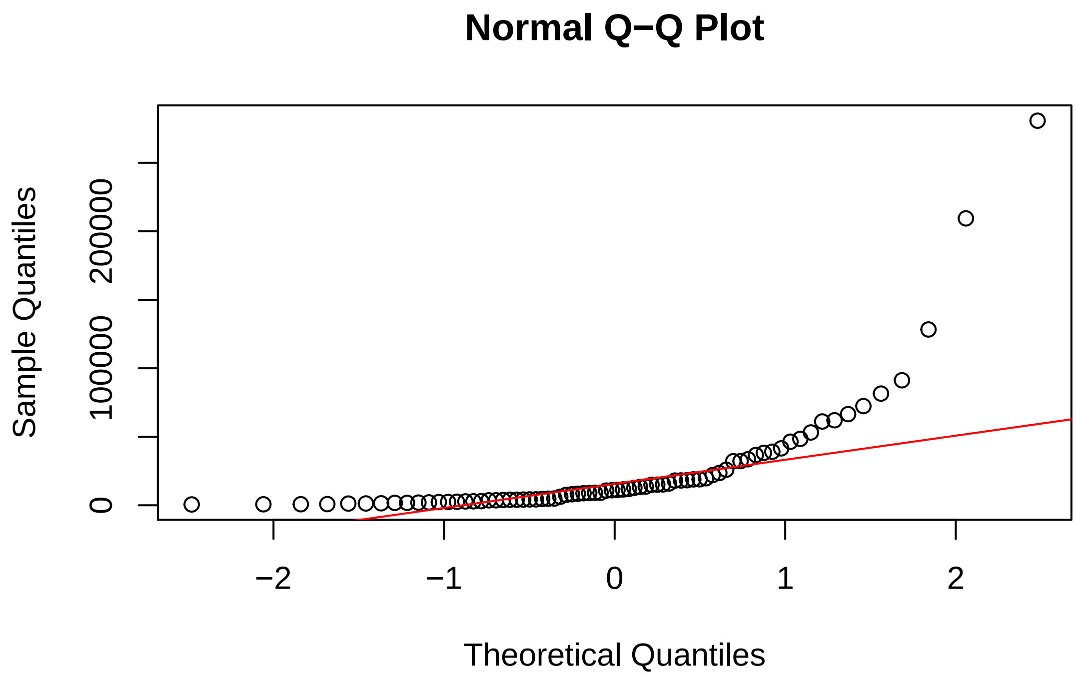
Questão 8)

Para os dois bancos de dados abaixo faça o qq-plot e a partir de um teste de hipotese de sua escolha (Shapiro-Wilk, Kolmogorov-Smirnov, entre outros) com 5% de significância, conclua se os dados possuem ou nao distribuição de probabilidade normal.

- O primeiro banco de dados e o "women_stem", que contém dados da seguinte reportagem: https://fivethirtyeight.com/features/the-economic-guide-to-picking-a-college-major/ que são dados de mulheres em trabalhos de ciência e tecnologia. Gostaríamos de verificar se os dados do Total de mulheres com majors em ciência e tecnologia possui distribuição de probabilidade normal.

```r
women_stem <- read.csv("women-stem.csv", sep=",", header = TRUE)

### Efetuando o teste da normalidade:
qqnorm(women_stem$Total);
qqline(women_stem$Total,col = 2)
```




```r
shapiro.test(women_stem$Total)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  women_stem$Total
## W = 0.54584, p-value = 5.728e-14
```

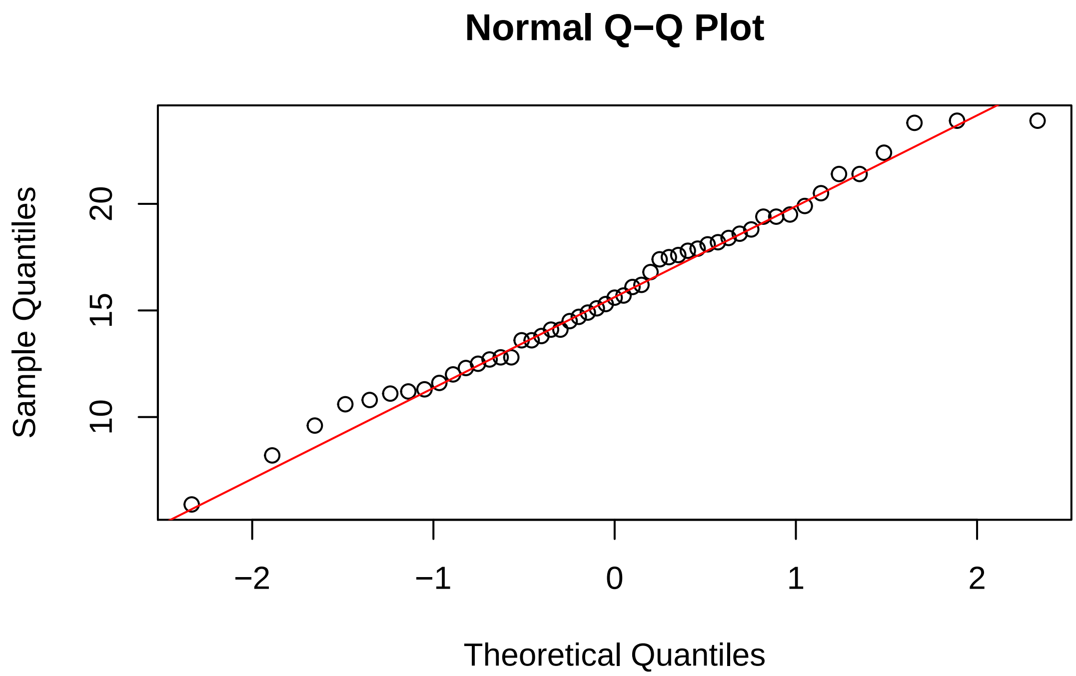
```r
ks.test(women_stem$Total, "pnorm", mean = mean(women_stem$Total), sd= sd(women_stem$Total))
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  women_stem$Total
## D = 0.28567, p-value = 5.54e-06
## alternative hypothesis: two-sided
```

```r
# Esses sao dados de contagem. Dessa forma, ja era de se esperar que os dados nao
# possuissem distribuicao normal. Pelo teste de hipoteses e qq-norm() corroboramos
# com essa ideia. Ou, os dados nao possuem distribuicao normal.
```

- O segundo banco de dados, e o mesmo utilizado na questao 5, sobre estados com os piores motoristas dos EUA. Dessa forma, verifique se a coluna que contém o número de motoristas envolvidos em colisões fatais por bilhão de milhas (Number.of.drivers.involved.in.fatal.collisions.per.billion.miles) possui distribuição normal.

```r
### Efetuando o teste da normalidade:
qqnorm(bad_drivers$Number.of.drivers.involved.in.fatal.collisions.per.billion.miles)
qqline(bad_drivers$Number.of.drivers.involved.in.fatal.collisions.per.billion.miles,
col = 2) #Maioria dos pontos proximos a linha de referencia->forte indicativo de normalidade
```




```r
variavel = bad_drivers$Number.of.drivers.involved.in.fatal.collisions.per.billion.miles
shapiro.test(variavel)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  variavel
## W = 0.98674, p-value = 0.8354
```

```r
ks.test(variavel, "pnorm", mean = mean(variavel), sd = sd(variavel))
```

```
## Warning in ks.test(variavel, "pnorm", mean = mean(variavel), sd = sd(variavel)):
## ties should not be present for the Kolmogorov-Smirnov test
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  variavel
## D = 0.063696, p-value = 0.9858
## alternative hypothesis: two-sided
```

```r
# p-valor > 0.05 -> temos evidencias amostrais ao nivel de significancia de 5%
# que a distribuicao de probabilidade da amostra e normal.
```

Questão 9)

Faça o exercício 1 pagina 141 do relatório técnico “BIOESTATÍSTICA BÁSICA USANDO O AMBIENTE COMPUTACIONAL R”, disponível no site do departamento de estatística da UFMG.

```r
sobrevivencia <- matrix(c(90,10,35,65),nrow = 2, byrow = TRUE,
                        dimnames = list(c("Seca", "Umida"), c("Sim","Não")))
```

```r
chisq.test(sobrevivencia)
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  sobrevivencia
## X-squared = 62.208, df = 1, p-value = 3.09e-15
```

```r
# Considerando 5%, temos evidencias amostrais que o efeito do inseticida e
# diferente entre as regioes seca e umida.
```

Questão 10)

Um pesquisador acredita que, numa determinada população, o número de descendentes deixados por indivíduo pode ser descrito por uma distribuição Poisson com $\lambda = 1$. A tabela abaixo apresenta as probabilidades calculadas para esta distribuição.

x	0	1	2	3	4	≥ 5
P(X=x)	0.3679	0.3679	0.1839	0.0613	0.0153	0.0037

Observando uma amostra de 500 pessoas desta população, o pesquisador encontrou os seguintes resultados, dados na tabela seguinte:

Número de filhos	Frequências observadas
0	170
1	180
2	95
3	35
4	18
≥ 5	2

O modelo de Poisson é adequado para descrever o número de descendentes deixados pelos indivíduos desta população? Considere nível de significância de 5%.

```r
 # obtendo probabilidades esperadas sob H0
p<-c(dpois(0:4,1), ppois(4,1,lower.tail=F) )
p
```

```
## [1] 0.367879441 0.367879441 0.183939721 0.061313240 0.015328310 0.003659847
```

```r
freqobs<-c(170,180,95,40,8,5)
chisq.test(x = freqobs,p = p)
```

```
## Warning in chisq.test(x = freqobs, p = p): Chi-squared approximation may be
## incorrect
```

```
##
##  Chi-squared test for given probabilities
##
## data:  freqobs
## X-squared = 9.6252, df = 5, p-value = 0.08658
```

```r
#Como a aproximacao Qui-Quadrado pode estar incorreta, vamos obter o P-valor por simulacao.
chisq.test(x=freqobs,p=p,simul=T)
```

```
##
##  Chi-squared test for given probabilities with simulated p-value (based
##  on 2000 replicates)
##
## data:  freqobs
## X-squared = 9.6252, df = NA, p-value = 0.08296
```

```r
# Nao ha evidencias amostrais de que o modelo proposto pelos pesquisadores para descrever
# o numero de descendentes seja inadequado, pois ao nível de significancia de 5%, nao rejeitamos H0.

# Esta questao é um exmeplo do uso do teste qui-quadrado para efetuar um teste de aderencia
```